

Amendments to the Specification

Please replace the title at page 1, line 1 and at page 34, line 1 with the following amended title:

**A DATA MINING SYSTEM FOR MINING INFORMATION ON PEOPLE AND
ORGANIZATIONS AND GENERATING BUSINESS E-MAIL ADDRESSES**

Please replace the paragraph at page 3, lines 5-10 with the following amended paragraph:

URL

URL stands for Uniform Resource Locator. Generally, URLs have three parts: the first part describes the protocol used to access the content pointed to by the URL, the second contains the directory in which the content is located, and the third contains the file that stores the content:

<protocol> : <domain> <directory> <file>

where "protocol" may be of the type http, "domain" is a domain name of the directory in which a file so named is located.

Please delete the paragraph at page 3, lines 11-15 which starts with "For example:".

Please replace the paragraph at page 3, lines 19-23 with the following amended paragraph:

For example, the following are legal variations of ~~the previous example~~ URLs:

www.corex.com/bios.html

www.cardscan.com

fn.cnn.com/archives/may99/pr37.html

ftp://shiva.lin.com/soft/words.zip

Please replace the paragraph at page 4, lines 18-26 with the following amended paragraph:

Decades of active research in the Computer Science field of Information Retrieval have ~~yield~~ yielded several algorithms and techniques for efficiently searching and retrieving information from structured databases. However, the world's largest information repository, the Web, contains mostly unstructured information, in the form of Web pages, text documents, or multimedia files. There are no standards on the content, format, or style of information published in the Web, except perhaps, the requirement that it should be understandable by human readers. Therefore the power of structured database queries that can readily connect, combine and filter information to present exactly what the user wants is not available in the Web.

Please replace the paragraph at page 16, lines 16-29 with the following amended paragraph:

To summarize, several Crawler 11 processes are needed by the system 40 in order to increase its efficiency, and an automated method must be employed to manage all these processes. The Distributor 47 offers exactly this functionality: it is a software module whose main function is to control and distribute work to multiple Crawlers 11. The Distributor 47 uses a database 14 to keep track of domain data 10 including all the Web sites that must be visited, and the visiting schedule for each one (some Web sites must be visited more frequently than others, depending on how often their contents change). In addition, the Distributor 47 prioritizes the Web sites according to their relative importance for the users, and it manages the Crawlers 11 so that the most important sites are visited first. The Distributor 40 is responsible to start multiple Crawler 11 processes, and keep their number as high as possible, without hurting the overall system performance. It also monitors the status of the running Crawler processes and stops or kills any processes that exhibit unwanted behavior (e.g. a process that takes too long, uses too much memory or disk space, etc).

Please replace the paragraph at page 18, lines 12-18 with the following amended paragraph:

For a detailed description of a preferred Extractor 41 that is customized to extract information about people from the Web see U.S. Patent Application No. 09/910,169, filed July 20, 2001 entitled "Computer Method and Apparatus for Extracting Data from Web

Pages", Attorney Docket No. 2937.1000-005. That Extractor 41 uses various methods and techniques described in U.S. Patent Application No. 09/585,320 filed on June 2, 2000 for a "Method and Apparatus for Deriving Information from Written Text".

Please replace the paragraph at page 21 lines 4-7 with the following amended paragraph:

A preferred embodiment of Loader 43 is described in the related U.S. Patent Application No. _____ 09/910,169, filed July 20, 2001, entitled "Computer Method and Apparatus for Extracting Data from Web Pages", Attorney Docket No. 2937.1000-005, cited above.